Madeleine Thompson
DSCI 551
Spring 2021

Midterm Progress Report

My project, Washington, D.C. Neighborhood Analysis, is coming along well. I have fully accessed my data set via the Yelp Fusion API and the Google Maps Distance Matrix API and have saved it in two csv files and a json file.

The below screenshot shows a snipped of zip_codes.csv. The data includes information on all zip codes in the states of DC, VA, and MD which encompass the area that I want to search over. I have saved information on the latitude and longitude, population, average house value, and average household income in each zip code. This data was scraped from https://www.zip-codes.com.

zip_codes

state	zip_code	zip_url	latitude	longitude	population	house_value	household_income
DC	20001	https://www.zip-codes.com/zip-code/20001/zip-code-20001.asp	38.910717	-77.01666	52586	659700	109318
DC	20002	https://www.zip-codes.com/zip-code/20002/zip-code-20002.asp	38.908089	-76.976663	76367	609300	84892
DC	20003	https://www.zip-codes.com/zip-code/20003/zip-code-20003.asp	38.877752	-76.98654600000000	38565	766900	126323
DC	20004	https://www.zip-codes.com/zip-code/20004/zip-code-20004.asp	38.892444	-77.020588	2219	546900	147522
DC	20005	https://www.zip-codes.com/zip-code/20005/zip-code-20005.asp	38.904224	-77.031798	13390	489900	94611
DC	20006	https://www.zip-codes.com/zip-code/20006/zip-code-20006.asp	38.895101	-77.039776	1137	231900	46319
DC	20007	https://www.zip-codes.com/zip-code/20007/zip-code-20007.asp	38.914063	-77.077395	27070	961100	130920
DC	20008	https://www.zip-codes.com/zip-code/20008/zip-code-20008.asp	38.935771	-77.059213	30244	776000	117195

The below screenshot shows a snippet of commute_info.csv. This data includes information on the location and distance of each zip code to my future place of employment in miles and as a commute time. This data was collected from the Google Maps Distance Matrix API - https://developers.google.com/maps.

commute_info

zip_code	latitude	longitude	distance	duration
20001	38.910717	-77.01666	6.1 mi	13 mins
20002	38.908089000000000	-76.976663	9.4 mi	25 mins
20003	38.877752	-76.986546	5.7 mi	13 mins
20004	38.892444	-77.020588	4.3 mi	9 mins
20005	38.904224	-77.031798	4.6 mi	12 mins
20006	38.895101000000000	-77.039776	3.0 mi	10 mins

The below screenshot shows a snippet of restaurant_data.json. This data includes information on every restaurant in each zip code scraped in the above zip_codes.csv file that includes “vegan”, “vegetarian”, or “cafe” in the description. This data was collected from the Yelp Fusion API - https://www.yelp.com/developers/documentation/v3/get_started.

I still need to work on saving my data to a database. I have downloaded and installed MongoDB and the process to store my existing data into a new database in MongoDB should be failry straightforward. Next, I need to look further into how I want to process the data in parallel and create a UI solution for my problem. I am having a harder time with both of these concepts because I have never attempted to parallel process data or create a UI before. I am on track and should be able to complete this project on time.